Faculty of Engineering and Information Technology
University of Technology Sydney

Online Media Information Enhancement by Fusing Multimodal Media Data

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

by

Lu Zhang

March 2021

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Lu Zhang declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Electrical and Data Engineering, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature: Signature removed prior to publication.

Date: 09/03/2021

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my principal supervisor, Associate Prof. Jian Zhang, for his professional guidance and help. He never gave up on me even when I stuck in my progress. I deeply appreciate my co-supervisor Jingsong Xu, for his help and guidance in both research and living in Sydney. He used to spend a lot of time to help me revise my paper and discuss my research problem. Also, I would like to thank Reader Jialie Shen for his guidance and advice in our collaborative research. Without his help, the research will not be completed successfully.

Besides, I want to thank my friend Yongshun Gong. Thanks for his patience and kindness. He always gave me generous help. I want to thank Zhibin Li. Thank him for his company and encouragement. Sometimes he was very strict with me and I believe he must have his reasons. I might not be able to finish my study without them. The time spent with them is my precious memory. I also want to thank myself. Thank you for being in a foreign country and supporting yourself with braveness and strength on countless nights. May the mountain remain the mountain. May the sea remain the sea. May we still be us.

Finally and most essentially, I would like to thank my parents. Without them, nothing would have any value.

Lu Zhang
March 2021 @ UTS

Contents

List of Figures

List of Tables

List of Publications

Papers Accepted and Published

- **Lu Zhang**, Jingsong Xu, Yongshun Gong, Litao Yu, Jian Zhang, and Jialie Shen, *Unsupervised Image and Text Fusion for Travel Information Enhancement*, IEEE Transactions on Multimedia, 2021.
- **Lu Zhang**, Jialie Shen, Jian Zhang, Jingsong Xu, Zhibin Li, Yazhou Yao, and Litao Yu, *Multimodal Marketing Intent Analysis for Effective Targeted Advertising*, IEEE Transactions on Multimedia, 2021.
- **Lu Zhang**, Jian Zhang, Jialie Shen, Jingsong Xu, Zhibin Li, and Litao Yu, *Incorporating Multimodal Cues for Advertorial Discovery*, International IEEE Conference on Multimedia and Expo, 2021.
- **Lu Zhang**, Jian Zhang, Zhibin Li, and Jingsong Xu, *Towards Better Graph Representation: Two-Branch Collaborative Graph Neural Networks for Multimodal Marketing Intention Detection*, International IEEE Conference on Multimedia and Expo, 2020.
- **Lu Zhang**, Jingsong Xu, Jian Zhang, and Yongshun Gong, *Information Enhancement for Travelogues via a Hybrid Clustering Model*, Digital Image Computing: Techniques and Applications, 2018.

Abstract

With the rapid improvement of human's ability to store and compute big data, more and more people are used to sharing and acquiring multimodal information on media platforms. The explosion of multimodal information is paving the way for many innovative applications based on the mining and analysis of big multimodal data. This thesis focuses on two applications: how to explore multimodal cues to recognise and analyse marketing intentions embedded in normal social news and how to mine correlative multimodal information for travel information understanding and recommendation.

Multimodal fusion is to join the associated features or integrate the intermediate decisions from two or more modalities in order to perform a prediction or an analysis for specific tasks. The key challenge behind this topic is the inherent heterogeneity of multimodal data leading to semantic gaps between them. Given a large amount of multimodal data, the main factor in establishing an effective model for specific tasks is the deep mining of inter-correlations between multiple modalities. Moreover, the intra-correlations within the same modality are also important for semantic inferring. This thesis develops several supervised and unsupervised methods to overcome these challenges.

To infer the commercial intentions embedded in multimodal content of social news, this thesis proposes a supervised Neural Autoregressive Topic Model which utilizes both textual words, visual words, and label information to learn discriminative representations for specific tasks. Moreover, this thesis introduces a supervised Two-branch Collaborative Graph Neural Net-

work, which uses a pre-defined fusion strategy to integrate news text and news images. A novel Multimodal Advertorial Discovery Model is proposed by adopting a Cross-graph Fusion strategy to achieve a soft assignment to incorporate images and text and generate comprehensive multimodal representations. For travelogues enhancement, this thesis introduces an unsupervised Hybrid Multiple Kernel K-means method to embed multimodal for matching travelogues and images. To verify the effectiveness of these proposed methods, experiments are conducted on real-life datasets from online media platforms.